Spectral Barriers in Certification Problems

Dmitriy (Tim) Kunisky — Dissertation Defense

April 27, 2021

I. Introduction

Our Question:

Are there efficient algorithms to certify bounds on random optimization problems?

Main example:

$$\mathsf{M}(\boldsymbol{W}) := \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}$$

(including Sherrington-Kirkpatrick and other Ising-type Hamiltonians, max-cut in graphs, synchronization over $\mathbb{Z}/2\mathbb{Z}$, etc.)

General *constrained PCA* problem:

$$\mathsf{M}_{\mathcal{X}}(\boldsymbol{W}) := \max_{\boldsymbol{X} \in \mathcal{X}} \mathsf{Tr}(\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})$$

(including Potts and vector-spin Hamiltonians; graph coloring; some CSPs; tensor, sparse, positive, conic, and other PCAs; etc.)

Search and Certification

Search and Certification

Search: Compute $\boldsymbol{x}_{\mathsf{alg}} = \boldsymbol{x}_{\mathsf{alg}}(\boldsymbol{W})$ with large value of objective $\boldsymbol{x}_{\mathsf{alg}}^{\top}\boldsymbol{W}\boldsymbol{x}_{\mathsf{alg}}$.

- Local ("greedy") search (or Markov chains, simulated annealing)
- Projected power method (or AMP variants)
- Relax and round

Search and Certification

Search: Compute $\boldsymbol{x}_{\mathsf{alg}} = \boldsymbol{x}_{\mathsf{alg}}(\boldsymbol{W})$ with large value of objective $\boldsymbol{x}_{\mathsf{alg}}^\top \boldsymbol{W} \boldsymbol{x}_{\mathsf{alg}}$.

- Local ("greedy") search (or Markov chains, simulated annealing)
- Projected power method (or AMP variants)
- Relax and round

Certification: Compute small $c(\boldsymbol{W}) \in \mathbb{R}$ that gives a bound $\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \leq c(\boldsymbol{W})$ *for all* feasible $\boldsymbol{x}$.

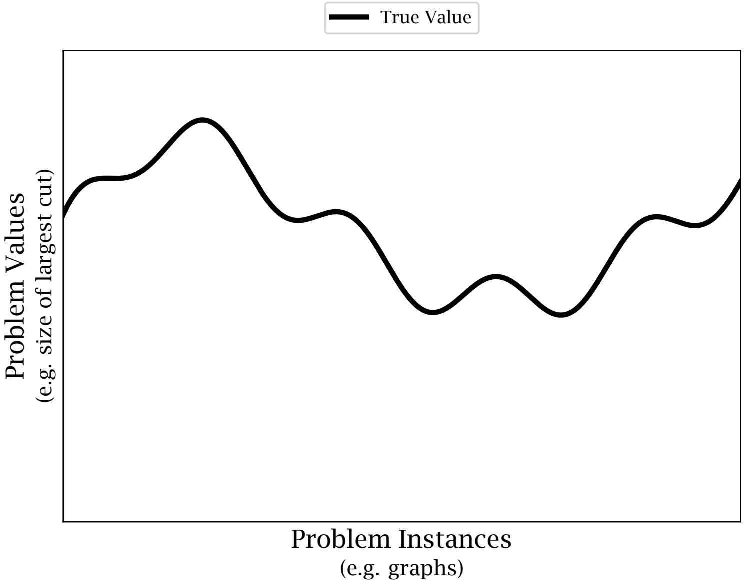
- LP relaxations (metric, Sherali-Adams, Lovász-Schrijver)
- SDP relaxations (Goemans-Williamson, **sum-of-squares/Lasserre**)

Search and Certification: Objective Bounds

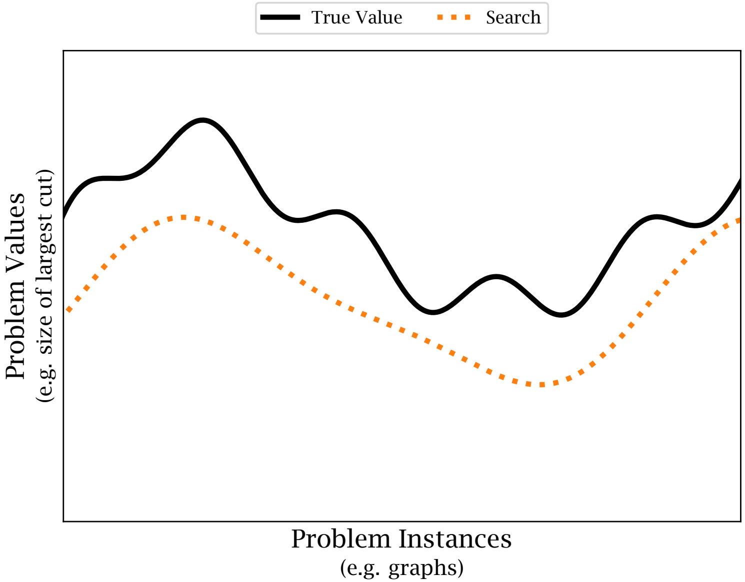
$$\boldsymbol{x}_{\mathsf{alg}}^{\top} \boldsymbol{W} \boldsymbol{x}_{\mathsf{alg}} \quad \leq \quad \mathsf{M}(\boldsymbol{W}) \quad \leq \quad c(\boldsymbol{W})$$

search	certification
$\updownarrow$	$\updownarrow$
lower bounds	**upper** bounds

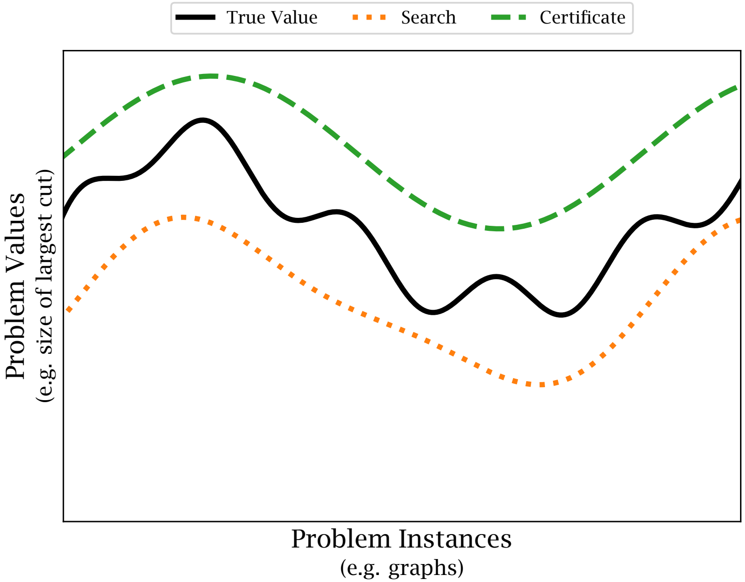
Average-Case Analysis



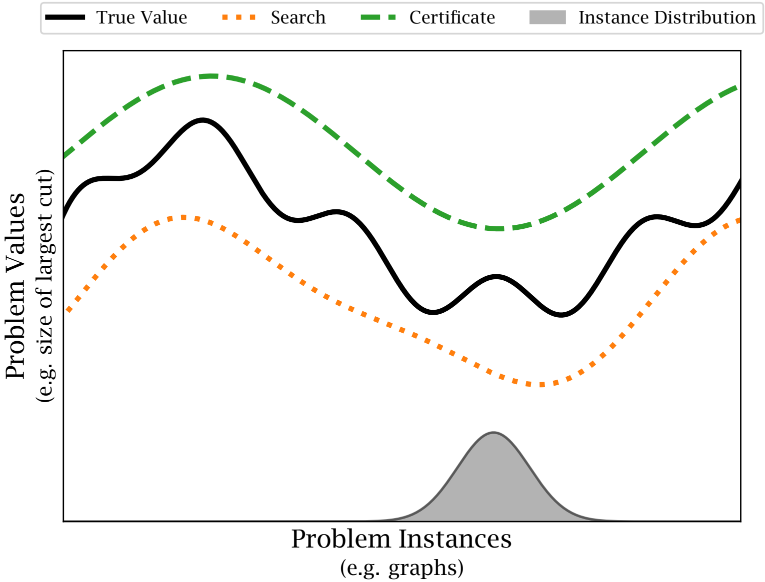
Average-Case Analysis



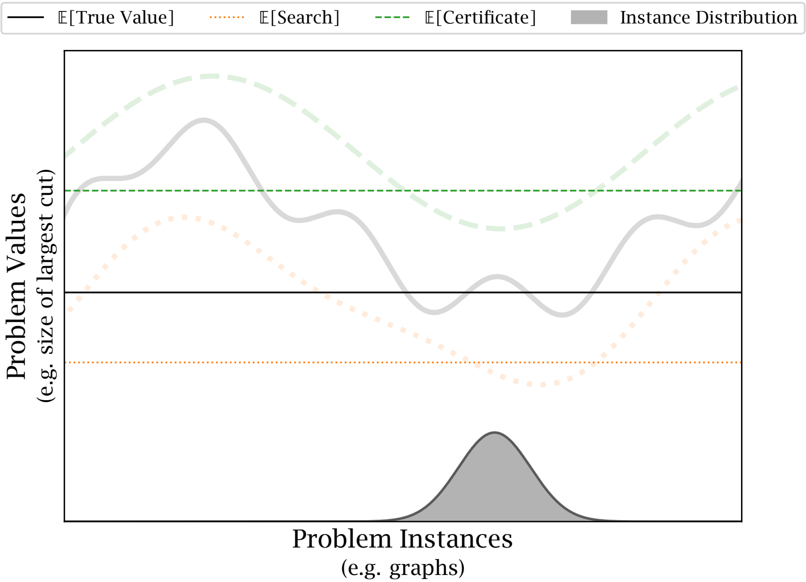
Average-Case Analysis



Average-Case Analysis



Average-Case Analysis



Gaussian Instance Distribution

Gaussian Instance Distribution

Take $\mathsf{M}(\boldsymbol{W}) = \mathsf{max}_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}$, for random $\boldsymbol{W} \sim \mathsf{GOE}(n)$.

Gaussian Instance Distribution

Take $\mathsf{M}(W) = \max_{x \in \{\pm 1\}^n} x^\top W x$, for random $W \sim \mathsf{GOE}(n)$.

Statistical physics tells us the true value:

$$\lim_{n \to \infty} \mathbb{E}\mathsf{M}(W)/n = 2\mathsf{P}_* \approx 1.526,$$ (“Parisi number”)

ground state of the SK model. [SK ’75; Parisi ’79; Guerra, Talagrand ’00s]

Gaussian Instance Distribution

Take $\mathsf{M}(W) = \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top W \boldsymbol{x}$, for random $W \sim \mathsf{GOE}(n)$.

Statistical physics tells us the true value:

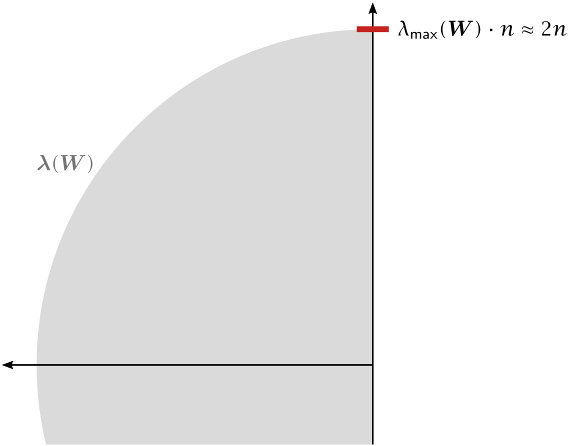
$$\lim_{n \to \infty} \mathbb{E}\mathsf{M}(\boldsymbol{W})/n = 2\mathsf{P}_* \approx 1.526,$$ ("Parisi number")

ground state of the SK model. [SK '75; Parisi '79; Guerra, Talagrand '00s]

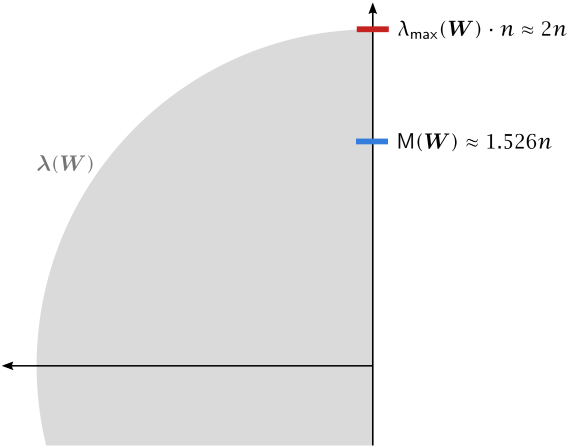
Other reasons to care:

- Natural sparse random graph limit
  [Boettcher, Zdeborová '10; Montanari, Sen '16; Dembo, Montanari, Sen '17]
- Proof complexity of ground state bounds (will see later)
- Mean width of cut polytope $\mathrm{conv}(\{\boldsymbol{x}\boldsymbol{x}^\top : \boldsymbol{x} \in \{\pm 1\}^n\})$

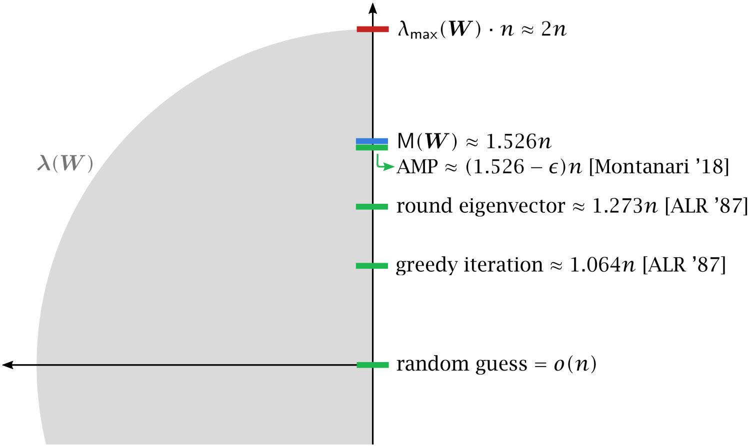
Searching for Hypercube in the GOE Spectrum



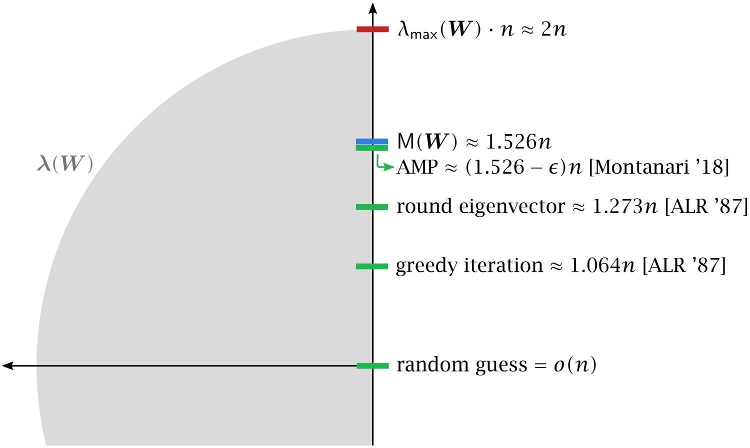
Searching for Hypercube in the GOE Spectrum



Searching for Hypercube in the GOE Spectrum



Searching for Hypercube in the GOE Spectrum



Optimal search! And certification?

Benchmark: The Spectral Certificate

Benchmark: The Spectral Certificate

The simplest way to bound $\mathsf{M}(\boldsymbol{W})$: ignore all special structure in $\boldsymbol{x}$, to get

$$\begin{aligned}
\mathsf{M}(\boldsymbol{W}) &= \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\
&\leq \max_{\|\boldsymbol{x}\| = \sqrt{n}} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\
&= \lambda_{\mathsf{max}}(\boldsymbol{W}) \cdot n \\
&\approx 2n, \qquad \text{(for } \boldsymbol{W} \sim \mathsf{GOE}(n)\text{)}
\end{aligned}$$

Benchmark: The Spectral Certificate

The simplest way to bound $\mathsf{M}(\boldsymbol{W})$: ignore all special structure in $\boldsymbol{x}$, to get

$$\begin{aligned}\mathsf{M}(\boldsymbol{W}) &= \max_{\boldsymbol{x}\in\{\pm1\}^n} \boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x} \\ &\leq \max_{\|\boldsymbol{x}\|=\sqrt{n}} \boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x} \\ &= \lambda_{\mathsf{max}}(\boldsymbol{W}) \cdot n \\ &\approx 2n, \qquad\qquad (\text{for } \boldsymbol{W} \sim \mathsf{GOE}(n))\end{aligned}$$

while a tight bound would give

$$\lesssim 1.526n.$$

Benchmark: The Spectral Certificate

The simplest way to bound $\mathsf{M}(\boldsymbol{W})$: ignore all special structure in $\boldsymbol{x}$, to get

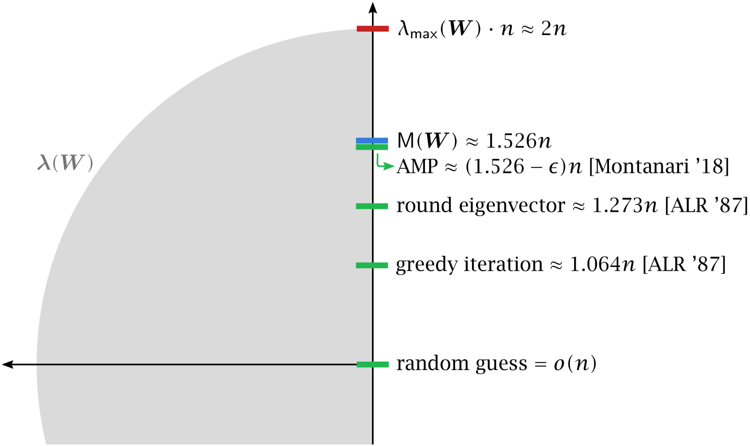
$$\begin{aligned}
\mathsf{M}(\boldsymbol{W}) &= \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\
&\leq \max_{\|\boldsymbol{x}\| = \sqrt{n}} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\
&= \lambda_{\max}(\boldsymbol{W}) \cdot n \\
&\approx 2n, && \text{(for } \boldsymbol{W} \sim \mathsf{GOE}(n)\text{)}
\end{aligned}$$

while a tight bound would give

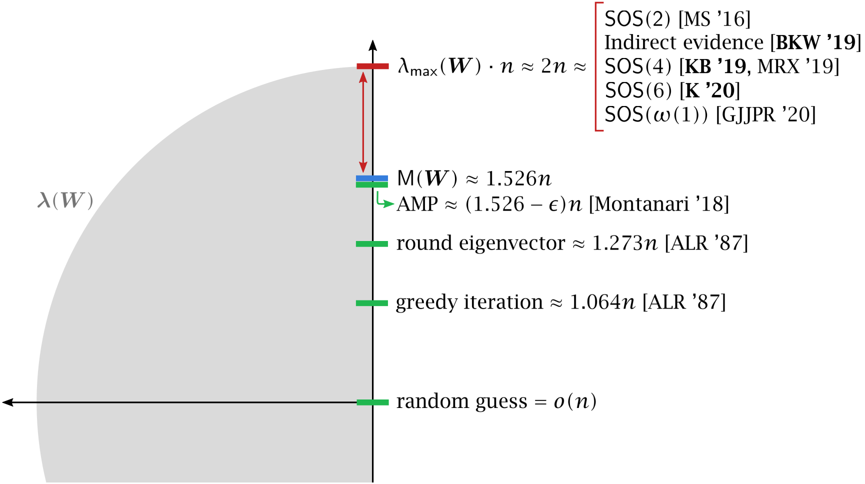
$$\lesssim 1.526n.$$

Can we do better?

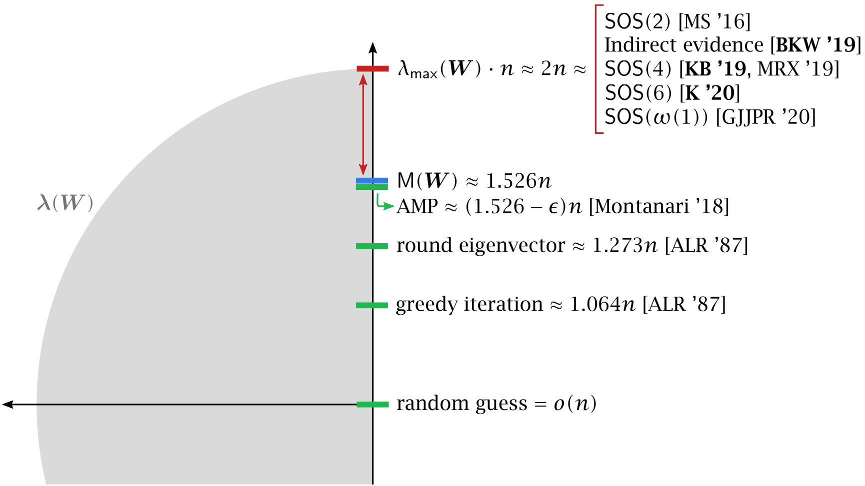
Certifying Hypercube in the GOE Spectrum



Certifying Hypercube in the GOE Spectrum



Certifying Hypercube in the GOE Spectrum



Optimal search, but **spectral barrier** to certification.

II. Computationally-Quiet Planting

Implicit Hypothesis Testing [WBP '16, BKW '19]

Implicit Hypothesis Testing [WBP '16, BKW '19]

Lemma (Reduction): Can certify $c(\boldsymbol{W}) \le (2 - \epsilon)n$ w.h.p. $\Rightarrow$ in same amount of time can test w.h.p. between:

1. $\boldsymbol{Y} \sim \mathbb{Q}_n$ (null model),
2. $\boldsymbol{Y} \sim \mathbb{P}_{n,\epsilon}$ (alternative/planted model),

models of eigenspaces of $\boldsymbol{W}$.

Implicit Hypothesis Testing [WBP '16, BKW '19]

Lemma (Reduction): Can certify $c(\boldsymbol{W}) \leq (2 - \epsilon)n$ w.h.p. $\Rightarrow$ in same amount of time can test w.h.p. between:

1. $\boldsymbol{Y} \sim \mathbb{Q}_n$ (null model),
2. $\boldsymbol{Y} \sim \mathbb{P}_{n,\epsilon}$ (alternative/planted model),

models of eigenspaces of $\boldsymbol{W}$.

Theorem (Hardness of Testing): The class of *low-degree polynomial algorithms* fails to distinguish $\mathbb{Q}_n$ from $\mathbb{P}_{n,\epsilon}$ for any $\epsilon > 0$.

Implicit Hypothesis Testing [WBP '16, BKW '19]

Lemma (Reduction): Can certify $c(\boldsymbol{W}) \leq (2-\epsilon)n$ w.h.p. $\Rightarrow$ in same amount of time can test w.h.p. between:

1. $\boldsymbol{Y} \sim \mathbb{Q}_n$ (null model),
2. $\boldsymbol{Y} \sim \mathbb{P}_{n,\epsilon}$ (alternative/planted model),

models of eigenspaces of $\boldsymbol{W}$.

Theorem (Hardness of Testing): The class of *low-degree polynomial algorithms* fails to distinguish $\mathbb{Q}_n$ from $\mathbb{P}_{n,\epsilon}$ for any $\epsilon > 0$.

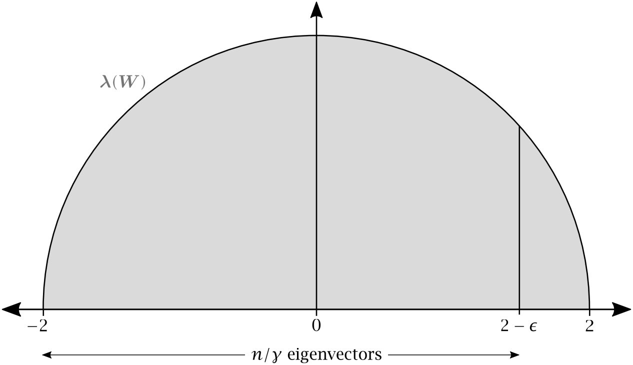
Corollary (Hardness of Certification): If low-degree polynomial algorithms are optimal tests, then no algorithm running in time $\exp(O(n^{1-\delta}))$ certifies $c(\boldsymbol{W}) \leq (2-\epsilon)n$.

Reduction from Spiked Matrix Model

Reduction from Spiked Matrix Model

GOE bottom eigenspaces vs. avoiding $\boldsymbol{x} \sim \mathsf{Unif}(\{\pm 1\}^n)$:

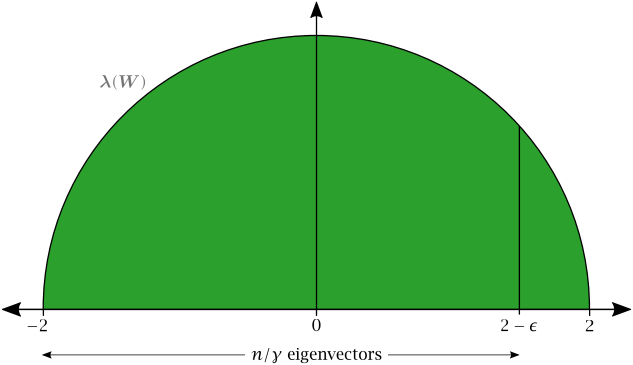
1. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \rightsquigarrow$ GOE,
2. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I} - \frac{\beta}{n}\boldsymbol{x}\boldsymbol{x}^\top) \rightsquigarrow$ spectrally-planted GOE.



Reduction from Spiked Matrix Model

GOE bottom eigenspaces vs. avoiding $\boldsymbol{x} \sim \mathsf{Unif}(\{\pm 1\}^n)$:

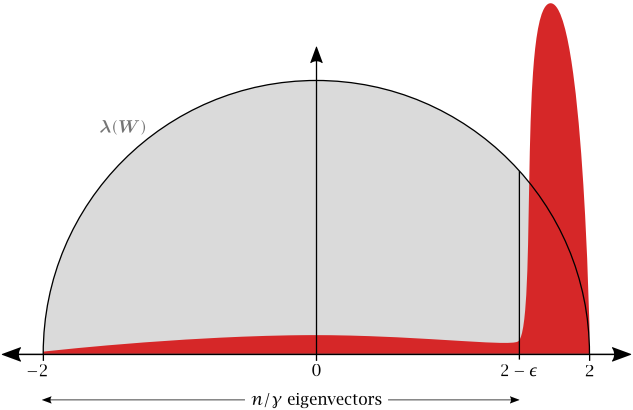
1. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \rightsquigarrow$ GOE,
2. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I} - \frac{\beta}{n}\boldsymbol{x}\boldsymbol{x}^\top) \rightsquigarrow$ spectrally-planted GOE.



Reduction from Spiked Matrix Model

GOE bottom eigenspaces vs. avoiding $\boldsymbol{x} \sim \mathsf{Unif}(\{\pm 1\}^n)$:

1. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \rightsquigarrow$ GOE,
2. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I} - \frac{\beta}{n}\boldsymbol{x}\boldsymbol{x}^\top) \rightsquigarrow$ spectrally-planted GOE.



Reduction from Spiked Matrix Model

GOE bottom eigenspaces vs. avoiding $\boldsymbol{x} \sim \mathsf{Unif}(\{\pm 1\}^n)$:

1. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \rightsquigarrow$ GOE,
2. $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\frac{n}{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I} - \frac{\beta}{n}\boldsymbol{x}\boldsymbol{x}^\top) \rightsquigarrow$ spectrally-planted GOE.

This is a **Wishart (negatively) spiked matrix model.**

[Johnstone '01, BBAP '05, BS '06, ...] but negative case first appreciated by [PWBM '18]

Gaussianity $\rightsquigarrow$ can do explicit calculations to assess difficulty of testing.

Similar versions for general setting: different or higher-rank constraints $\mathcal{X}$ with random $\boldsymbol{X}$ near $\mathcal{X}$.

[BBKMW '20, BKW '20, Chapter 2 of thesis]

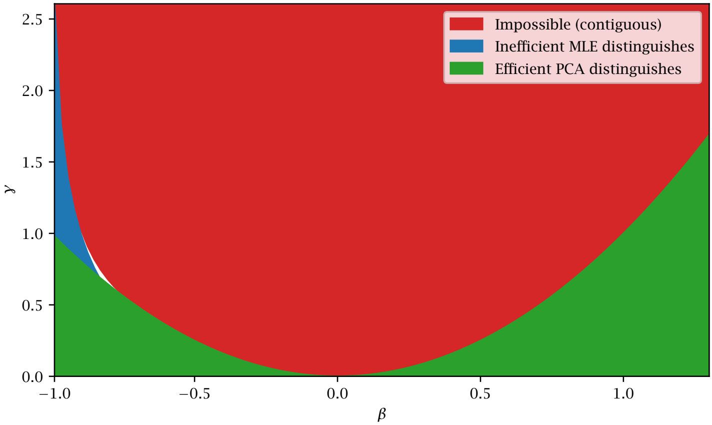
Hard Regime for Testing? [BS ’06, PWBM ’18]

Hard Regime for Testing? [BS ’06, PWBM ’18]

Natural test: “PCA” or threshold $\lambda_{\mathsf{min}|\mathsf{max}}(\sum \boldsymbol{y}_i \boldsymbol{y}_i^\top)$.

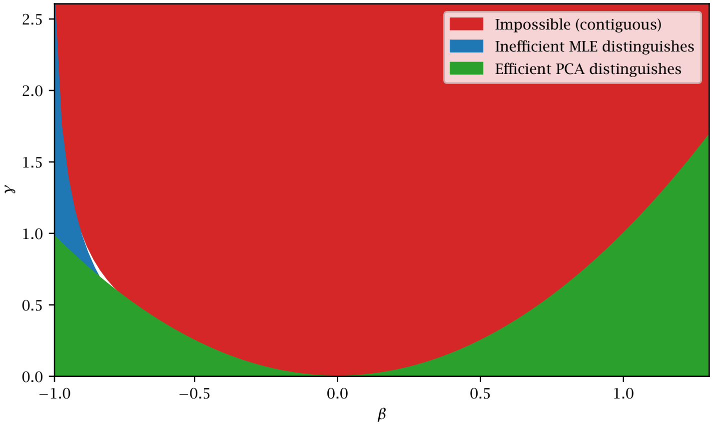
Hard Regime for Testing? [BS '06, PWBM '18]

Natural test: "PCA" or threshold $\lambda_{\mathsf{min}|\mathsf{max}}(\sum \boldsymbol{y}_i \boldsymbol{y}_i^\top)$.



Hard Regime for Testing? [BS ’06, PWBM ’18]

Natural test: “PCA” or threshold $\lambda_{\mathsf{min}|\mathsf{max}}(\sum \boldsymbol{y}_i \boldsymbol{y}_i^\top)$.



Better evidence of hardness than just PCA failure?

Low-Degree Tests [HKPRSS '17, HS '17, Hopkins '18]

Low-Degree Tests [HKPRSS '17, HS '17, Hopkins '18]

Take {low-degree polynomials} ≈ {efficient algorithms}, and compute:

$$\begin{aligned} \text{maximize} \quad & \mathbb{E}_{\mathbb{P}_n}\, p(\boldsymbol{Y}) \\ \text{subject to} \quad & \mathbb{E}_{\mathbb{Q}_n} p(\boldsymbol{Y})^2 \le 1, \\ & p(\boldsymbol{Y}) \in \mathbb{R}[\boldsymbol{Y}]_{\le D}. \end{aligned}$$

Low-Degree Tests [HKPRSS ’17, HS ’17, Hopkins ’18]

Take {low-degree polynomials} ≈ {efficient algorithms}, and compute:

$$\begin{aligned} \text{maximize} \quad & \left\langle p, \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\rangle \\ \text{subject to} \quad & \|p\|^2 \leq 1, \\ & p \in \mathbb{R}[\boldsymbol{Y}]_{\leq D} \subset L^2(\mathbb{Q}_n). \end{aligned}$$

Low-Degree Tests [HKPRSS '17, HS '17, Hopkins '18]

Take {low-degree polynomials} ≈ {efficient algorithms}, and compute:

$$\begin{aligned} \text{maximize} \quad & \left\langle p, \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\rangle \\ \text{subject to} \quad & \|p\|^2 \le 1, \\ & p \in \mathbb{R}[\boldsymbol{Y}]_{\le D} \subset L^2(\mathbb{Q}_n). \end{aligned}$$

Optimizer: the (normalized) **low-degree likelihood ratio**

$$p^\star(\boldsymbol{Y}) = \mathcal{P}^{\le D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}) \;\Big/\; \underbrace{\left\| \mathcal{P}^{\le D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|}_{\text{objective value}}.$$

Low-Degree Tests [HKPRSS '17, HS '17, Hopkins '18]

Take {low-degree polynomials} ≈ {efficient algorithms}, and compute:

$$\begin{aligned} \text{maximize} \quad & \left\langle p, \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\rangle \\ \text{subject to} \quad & \|p\|^2 \le 1, \\ & p \in \mathbb{R}[\boldsymbol{Y}]_{\le D} \subset L^2(\mathbb{Q}_n). \end{aligned}$$

Optimizer: the (normalized) **low-degree** **likelihood ratio**

$$p^\star(\boldsymbol{Y}) = \mathcal{P}^{\le D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}) \;\; / \;\; \underbrace{\left\| \mathcal{P}^{\le D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|}_{\text{objective value}}.$$

Conjecture: Cannot test in time $e^{\widetilde{O}(D(n))}$ if objective $= O(1)$.

Low-Degree Lower Bounds

Low-Degree Lower Bounds

New evidence for hard or subexponential regimes in:

- Dense matrix PCA (Wigner/**Wishart**, rank k, many priors) [BKW ’19]
- One special non-Gaussian dense matrix PCA [K ’20]
- Sparse matrix PCA (rank 1, many priors) [DKWB ’19]
- Tensor PCA (many priors) [KWB ’19]
- Stochastic block model [BBKMW ’20]

Low-Degree Lower Bounds

New evidence for hard or subexponential regimes in:

- Dense matrix PCA (Wigner/**Wishart**, rank k, many priors) [BKW '19]
- One special non-Gaussian dense matrix PCA [K '20]
- Sparse matrix PCA (rank 1, many priors) [DKWB '19]
- Tensor PCA (many priors) [KWB '19]
- Stochastic block model [BBKMW '20]

From Wishart models + reduction, conclude conditional hardness of better-than-spectral certification in:

- SK Hamiltonian ("Gaussian max-cut") [BKW '19]
- Potts glass Hamiltonian ("Gaussian coloring") [BBKMW '20]
- Positive PCA ("Gaussian max-clique") [BKW '20]

Unified Proofs from Overlap Formulae

Unified Proofs from Overlap Formulae

Algebraic simplification of orthogonal polynomials method: when in planted $\mathbb{P}_n$ we observe noisy signal $\widetilde{\boldsymbol{X}}$ (e.g. $= \boldsymbol{x}\boldsymbol{x}^\top$),

$$\left\| \mathcal{P}^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \sum_{|\boldsymbol{\alpha}| \leq D} \left\langle q_{\boldsymbol{\alpha}}, \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\rangle^2 \leq \mathop{\mathbb{E}}_{\widetilde{\boldsymbol{X}}^1, \widetilde{\boldsymbol{X}}^2} \sum_{d=0}^{D} c_d \, r(\widetilde{\boldsymbol{X}}^1, \widetilde{\boldsymbol{X}}^2)^d$$

Unified Proofs from Overlap Formulae

Algebraic simplification of orthogonal polynomials method: when in planted $\mathbb{P}_n$ we observe noisy signal $\tilde{\boldsymbol{X}}$ (e.g. $= \boldsymbol{x}\boldsymbol{x}^\top$),

$$\left\| \mathcal{P}^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \sum_{|\boldsymbol{\alpha}| \leq D} \left\langle q_{\boldsymbol{\alpha}}, \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\rangle^2 \leq \mathop{\mathbb{E}}_{\tilde{\boldsymbol{X}}^1, \tilde{\boldsymbol{X}}^2} \sum_{d=0}^{D} c_d r(\tilde{\boldsymbol{X}}^1, \tilde{\boldsymbol{X}}^2)^d$$

Reduce to contest of low-dimensional **replica overlap** $r(\tilde{\boldsymbol{X}}^1, \tilde{\boldsymbol{X}}^2)$ tails vs. **link function** $f(t) = \sum_{d=0}^{\infty} c_d t^d$ growth.

Unified Proofs from Overlap Formulae

Algebraic simplification of orthogonal polynomials method: when in planted $\mathbb{P}_n$ we observe noisy signal $\widetilde{\boldsymbol{X}}$ (e.g. $= \boldsymbol{x}\boldsymbol{x}^\top$),

$$\left\|\mathcal{P}^{\leq D}\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right\|^2 = \sum_{|\boldsymbol{\alpha}|\leq D}\left\langle q_{\boldsymbol{\alpha}}, \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right\rangle^2 \leq \mathop{\mathbb{E}}_{\widetilde{\boldsymbol{X}}^1,\widetilde{\boldsymbol{X}}^2}\sum_{d=0}^{D} c_d\, r(\widetilde{\boldsymbol{X}}^1,\widetilde{\boldsymbol{X}}^2)^d$$

Reduce to contest of low-dimensional **replica overlap** $r(\widetilde{\boldsymbol{X}}^1,\widetilde{\boldsymbol{X}}^2)$ tails vs. **link function** $f(t) = \sum_{d=0}^{\infty} c_d t^d$ growth.

Model	**Overlap**	**Link Function**
Gaussian Wigner	$\langle\widetilde{\boldsymbol{X}}^1,\widetilde{\boldsymbol{X}}^2\rangle$	$\exp(t)$
Morris Exp. Families	$\langle\boldsymbol{z}(\widetilde{\boldsymbol{X}}^1),\boldsymbol{z}(\widetilde{\boldsymbol{X}}^2)\rangle$	$(1-vt)^{-1/v}$
Gaussian Wishart	$\widetilde{\boldsymbol{X}}^1\widetilde{\boldsymbol{X}}^2$	$\det(\boldsymbol{I}-\boldsymbol{T})^{-n/2\gamma}$

III. Sum-of-Squares Lower Bounds

State of Affairs So Far

State of Affairs So Far

Reduction + low-degree analysis $\rightsquigarrow$ evidence of no efficient better-than-spectral certification algorithm.

What more could we want?

State of Affairs So Far

Reduction + low-degree analysis $\rightsquigarrow$ evidence of no efficient better-than-spectral certification algorithm.

What more could we want?

1. $\boldsymbol{W} \sim$ something other than $\mathsf{GOE}(n)$; in particular, analysis **without exact law of eigenvectors**.

State of Affairs So Far

Reduction + low-degree analysis $\rightsquigarrow$ evidence of no efficient better-than-spectral certification algorithm.

What more could we want?

1. $\boldsymbol{W} \sim$ something other than $\mathsf{GOE}(n)$; in particular, analysis **without exact law of eigenvectors**.

2. **Concrete lower bounds** rather than “evidence” conditional on conjecture re: low-degree polynomials.

Sum-of-Squares Relaxations [Parrilo '00, Lasserre '01]

A way to design increasingly powerful **semidefinite programming** relaxations of $\mathsf{M}(\boldsymbol{W}) = \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}$.

Sum-of-Squares Relaxations [Parrilo '00, Lasserre '01]

A way to design increasingly powerful **semidefinite programming** relaxations of $\mathsf{M}(\boldsymbol{W}) = \max_{\boldsymbol{x}\in\{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}$.

$$
\begin{aligned}
\max \quad & \widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] \\
\text{s.t.} \quad & \widetilde{\mathbb{E}} : \mathbb{R}[x_1, \dots, x_n]_{\leq D} \to \mathbb{R} \\
& \widetilde{\mathbb{E}} \text{ linear} \\
& \widetilde{\mathbb{E}}[1] = 1 \\
& \widetilde{\mathbb{E}}[(x_i^2 - 1) p(x)] = 0 \\
& \widetilde{\mathbb{E}}[s(x)^2] \geq 0
\end{aligned}
$$

pseudoexpectation,
imposter,
charlatan,
...

Sum-of-Squares Relaxations [Parrilo '00, Lasserre '01]

A way to design increasingly powerful **semidefinite programming** relaxations of $\mathsf{M}(\boldsymbol{W}) = \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}$.

$$\begin{array}{ll} \max & \widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] \\ \text{s.t.} & \widetilde{\mathbb{E}} : \mathbb{R}[x_1, \dots, x_n]_{\leq D} \to \mathbb{R} \\ & \widetilde{\mathbb{E}} \text{ linear} \\ & \widetilde{\mathbb{E}}[1] = 1 \\ & \widetilde{\mathbb{E}}[(x_i^2 - 1) p(x)] = 0 \\ & \widetilde{\mathbb{E}}[s(x)^2] \geq 0 \end{array} = \begin{array}{ll} \min & c \\ \text{s.t.} & c - \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\ & \quad = \sum_i p_i(x)(x_i^2 - 1) \\ & \quad + \sum_j s_j(x)^2 \\ & \deg p_i \leq D - 2 \\ & \deg s_j \leq D/2 \end{array}$$

pseudoexpectation,
imposter,
charlatan,
...

sum-of-squares prover,
detective,
tester,
...

Sum-of-Squares Relaxations [Parrilo '00, Lasserre '01]

A way to design increasingly powerful **semidefinite programming** relaxations of $\mathsf{M}(\boldsymbol{W}) = \max_{\boldsymbol{x} \in \{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}$.

$$
\begin{array}{ll}
\max & \widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] \\
\text{s.t.} & \widetilde{\mathbb{E}} : \mathbb{R}[x_1, \dots, x_n]_{\leq D} \to \mathbb{R} \\
& \widetilde{\mathbb{E}} \text{ linear} \\
& \widetilde{\mathbb{E}}[1] = 1 \\
& \widetilde{\mathbb{E}}[(x_i^2 - 1) p(x)] = 0 \\
& \widetilde{\mathbb{E}}[s(x)^2] \geq 0
\end{array}
\quad = \quad
\begin{array}{ll}
\min & c \\
\text{s.t.} & c - \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\
& \quad = \sum_i p_i(x)(x_i^2 - 1) \\
& \quad + \sum_j s_j(x)^2 \\
& \deg p_i \leq D - 2 \\
& \deg s_j \leq D/2
\end{array}
$$

pseudoexpectation, imposter, charlatan, ...

sum-of-squares prover, detective, tester, ...

Pay runtime of $n^{\Theta(D)}$ for flexibility of degree D proofs.

Pseudocalibration [MW '13...BHKKMP '16]

Pseudocalibration [MW '13...BHKKMP '16]

To give a lower bound, need $\widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] = \langle \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top], \boldsymbol{W} \rangle \approx 2n$.

Pseudocalibration [MW '13...BHKKMP '16]

To give a lower bound, need $\widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] = \langle \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top], \boldsymbol{W} \rangle \approx 2n$.

Choosing $\widetilde{\mathbb{E}}$ = choosing each $\widetilde{\mathbb{E}}[x_{i_1} \cdots x_{i_D}]$, starting with $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] \propto$ projector to top eigenspace V of $\boldsymbol{W}$. ($D = 2$ [MS '16])

Pseudocalibration [MW '13...BHKKMP '16]

To give a lower bound, need $\tilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] = \langle \tilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top], \boldsymbol{W} \rangle \approx 2n$.

Choosing $\tilde{\mathbb{E}}$ = choosing each $\tilde{\mathbb{E}}[x_{i_1} \cdots x_{i_D}]$, starting with $\tilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] \propto$ projector to top eigenspace V of $\boldsymbol{W}$. ($D = 2$ [MS '16])

Use planted distribution ("$\boldsymbol{x} \in V \cap \{\pm 1\}^n$") to determine these, entry by entry. Only difficulty: **proving positivity**.

Pseudocalibration [MW '13...BHKKMP '16]

To give a lower bound, need $\widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x}] = \langle \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top], \boldsymbol{W} \rangle \approx 2n$.

Choosing $\widetilde{\mathbb{E}}$ = choosing each $\widetilde{\mathbb{E}}[x_{i_1} \cdots x_{i_D}]$, starting with $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] \propto$ projector to top eigenspace V of $\boldsymbol{W}$. ($D = 2$ [MS '16])

Use planted distribution ("$\boldsymbol{x} \in V \cap \{\pm 1\}^n$") to determine these, entry by entry. Only difficulty: **proving positivity**.

Question: Whose **Gram matrices** are these large structured psd matrices? How to make more explicit these **geometric** objects underlying SOS lower bounds?

Spectral Extensions [BK '18, KB '19, K '20]

Spectral Extensions [BK '18, KB '19, K '20]

Probabilistic viewpoint: use *surrogate random tensors*,

$$\widetilde{\mathbb{E}}[\boldsymbol{x^i x^j}] := \mathbb{E}[G_{\boldsymbol{i}}^{(d)} G_{\boldsymbol{j}}^{(d)}] \quad \leftrightarrow \quad \text{“}\boldsymbol{G}^{(d)} = \boldsymbol{x}^{\otimes d}\text{”}$$

Spectral Extensions [BK '18, KB '19, K '20]

Probabilistic viewpoint: use *surrogate random tensors*,

$$\widetilde{\mathbb{E}}[\boldsymbol{x}^{\boldsymbol{i}}\boldsymbol{x}^{\boldsymbol{j}}] := \mathbb{E}[G_{\boldsymbol{i}}^{(d)} G_{\boldsymbol{j}}^{(d)}] \quad \leftrightarrow \quad \text{“}\boldsymbol{G}^{(d)} = \boldsymbol{x}^{\otimes d}\text{”}$$

Make $\boldsymbol{G}^{(d)}$ as random as possible under constraints—give it a canonical Gaussian distribution conditional on:

$$G_{\emptyset}^{(0)} = 1 \quad \leftrightarrow \quad x^{\otimes 0} = (1)$$

$$\boldsymbol{G}^{(d)} \text{ symmetric} \quad \leftrightarrow \quad x_{i_1} \cdots x_{i_d} = x_{\pi(i_1)} \cdots x_{\pi(i_d)}$$

$$G_{j,j,i_1,\dots,i_{d-2}}^{(d)} = G_{i_1,\dots,i_{d-2}}^{(d-2)} \quad \leftrightarrow \quad x_j^2 \boldsymbol{x}^{\boldsymbol{i}} = \boldsymbol{x}^{\boldsymbol{i}}$$

$$(G_{j,i_1,\dots,i_{d-1}}^{(d)})_{j=1}^{n} \in V \quad \leftrightarrow \quad x_{i_1} \cdots x_{i_{d-1}} \boldsymbol{x} \in V$$

for V top eigenspace of $\boldsymbol{W}$.

The Surprises of Spectral Extensions

The Surprises of Spectral Extensions

With homogeneous polynomials $\rightsquigarrow$ ideal generated by $\langle \boldsymbol{P}_V \boldsymbol{e}_i, \boldsymbol{z} \rangle^2$ and *multiharmonic* polynomials $\langle \boldsymbol{P}_V \boldsymbol{e}_i, \boldsymbol{\partial} \rangle^2 p = 0$.

The Surprises of Spectral Extensions

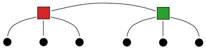
With homogeneous polynomials $\rightsquigarrow$ ideal generated by $\langle \boldsymbol{P}_V \boldsymbol{e}_i, \boldsymbol{z} \rangle^2$ and *multiharmonic* polynomials $\langle \boldsymbol{P}_V \boldsymbol{e}_i, \boldsymbol{\partial} \rangle^2 p = 0$.

Multiharmonic projections $\rightsquigarrow$ Green's function representation gives diagrammatic expressions. [Maxwell 1873!]

$\widetilde{\mathbb{E}}[x_i x_j x_k x_\ell x_m x_p]$ = sum of expressions of the form

$P_{ij} P_{k\ell} P_{mp}$

$P_{ij} \sum_{a=1}^{n} P_{ak} P_{a\ell} P_{am} P_{ap}$



$\sum_{a,b=1}^{n} P_{ai} P_{aj} P_{ak} P_{ab} P_{b\ell} P_{bm} P_{bp}$

The Surprises of Spectral Extensions

With homogeneous polynomials $\rightsquigarrow$ ideal generated by $\langle \boldsymbol{P}_V \boldsymbol{e}_i, \boldsymbol{z} \rangle^2$ and *multiharmonic* polynomials $\langle \boldsymbol{P}_V \boldsymbol{e}_i, \boldsymbol{\partial} \rangle^2 p = 0$.

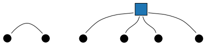
Multiharmonic projections $\rightsquigarrow$ Green's function representation gives diagrammatic expressions. [Maxwell 1873!]

Möbius function coefficients $\rightsquigarrow$ identities "explaining" compatibility of positivity and entry symmetries.

$$\widetilde{\mathbb{E}}[x_i x_j x_k x_\ell x_m x_p] = \text{sum of expressions of the form}$$

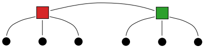
$P_{ij}P_{k\ell}P_{mp}$

1



$P_{ij} \sum_{a=1}^{n} P_{ak}P_{a\ell}P_{am}P_{ap}$

−2



$\sum_{a,b=1}^{n} P_{ai}P_{aj}P_{ak}P_{ab}P_{b\ell}P_{bm}P_{bp}$

4

Results on Sum-of-Squares

Results on Sum-of-Squares

With spectral extensions:

- Deterministic analysis of degree 4 feasible set [BK '18]
- Degree 4 lower bound for SK Hamiltonian [KB '19]
- Towards higher-degree lower bounds: [K '20]
  - Degree 6 lower bound for SK Hamiltonian
  - Degree $\omega(1)$ for incoherent high-rank projector
- Spectrum of parity $\widetilde{\mathbb{E}}$ [Grigoriev '01, Laurent '03, BKM '21]

Results on Sum-of-Squares

With spectral extensions:

- Deterministic analysis of degree 4 feasible set [BK '18]
- Degree 4 lower bound for SK Hamiltonian [KB '19]
- Towards higher-degree lower bounds: [K '20]
  - Degree 6 lower bound for SK Hamiltonian
  - Degree $\omega(1)$ for incoherent high-rank projector
- Spectrum of parity $\widetilde{\mathbb{E}}$ [Grigoriev '01, Laurent '03, BKM '21]

Meanwhile, with pseudocalibration:

- Degree 4 lower bound for SK Hamiltonian [MRX '19]
- Degree $\omega(1)$ lower bound for SK Hamiltonian [GJJPR '20]

Some Future Directions

Some Future Directions

- Low-degree method beyond “integrable” models
- Connection with statistical physics via overlaps
- Combinatorial manifestations of spectral planting

Some Future Directions

- Low-degree method beyond “integrable” models
- Connection with statistical physics via overlaps
- Combinatorial manifestations of spectral planting

- Spectra and Gramian structure of pseudoexpectations
- Pseudocalibration reconciliation
- Proof systems for spin glasses—replicated SOS? [RS ’00]

Thank you!